Application of Artificial Neural Networks for Reduction of False-Positive Detections in Digital Chest Radiographs

Jyh-Shyan Lin*,†, Panos A. Ligomenides*, Matthew T. Freedman†, and Seong K. Mun†

*Cybernetics Research Laboratory, Electrical Engineering Department,
University of Maryland, College Park, MD 20742
and
†ISIS Center, Radiology Department,
Georgetown University Medical Center, Washington, D.C. 20007

ABSTRACT

A methodology based on the fuzzy set theory and the convolution neural network (CNN) architecture is proposed to tackle the problem of reducing false-positive rate in automatic lung nodule detection [1]. The CNN which simulates human visual mechanism was trained by a supervised back-propagation algorithm based on fuzzy membership functions [2]. The training and testing database consists of image blocks (each 32 × 32 pixels) of suspected lung nodule areas (nodule candidates) which were generated from our pre-scanning program [1]. A linguistic label was assigned to each nodule candidate of the training set, then the label was converted to a membership value through a pre-defined membership function [3] and used as teaching signal (desired outputs) during the network learning. Before the nodule candidate was fed to the network input, it was pre-processed to reduce the complex background noise and the contrast discrepancy resulted from film development. During the network testing phase , a defuzzification process [2] was applied to decipher the trained network's output triggered by the nodule candidate in the testing set. Finally, a Receiver Operating Characteristic (ROC) analysis [3] was used to evaluate the CNN's performance based on the defuzzified output of the testing database. Preliminary results showed an average A_Z (the performance index) of 0.84 which is equivalent to 0.80 true-positive detection (sensitivity) with an average 2 ~ 3 false-positive detections per chest image .

I. INTRODUCTION

Lung cancer is one of the most common and deadly diseases in the world. The cure of lung cancer depends highly on the early detection and treatment of small and localized tumors. As reported by Heelan [4], the detection of lung tumors in the early stage of growth can result in a better prognosis for survival. The detection and diagnosis of pulmonary nodules in chest radiographs are among the most difficult clinical tasks performed by radiologists. Due to the human observer errors, currently the miss rate in detecting lung nodules (size from 3 mm to 20 mm in diameter) is as high as 35% of the abnormal cases, of which one third of the missed nodules can be detected retrospectively. However, the miss rate can be decreased (< 20%) if two or more radiologists work together. Among the researches of improving the diagnostic accuracy [5][6][7][8], computer-assisted diagnosis (CAD) has been concluded as a promising approach.

Recently, various CAD schemes, which utilize both digital image processing techniques and artificial neural networks (ANNs), have been enthusiastically proposed to assist radiologists in detecting of lung nodules [9][10][11]. The CAD schemes perform mainly two diagnostic functions: (1) locating the suspected nodule areas (pre-scan process) and (2) differentiating the true nodules from the false nodules. Although many digital processing algorithms for locating and differentiating suspected nodules have been proposed [5][6], too many false-positive detections were reported. Therefore, our goal is to apply advanced ANN technologies to the CAD scheme to reduce the number of false-positive detections while maintaining a high true-positive detection rate.

For distinguishing true and false nodules, we are investigating a CNN architecture. The CNN network can synthesize appropriate image feature extractors through a supervised learning process (to be described in Section II). Our database consists of nodule candidates (the 32 × 32 pixel image blocks of suspected nodule areas) generated from the pre-scan process [1]. Each nodule candidate was assigned a linguistic label, such as "definitely a nodule", "probably a nodule", "possibly a nodule", "probably no nodule", or "definitely no nodule", by a radiologist. The labels which reflect the radiologist's diagnoses were fuzzified [2] and used as teaching signals to supervise the network learning. During the network testing, the CNN's outputs were first defuzzified [2] and then analyzed by using the ROC method.

In our computer simulation, we used genuine nodules instead of simulated nodules [9][12] to test the network's ability in dealing with the problem of lung nodule detection in digital chest radiographs.

0195-4210/92/$5.00

II. MATERIALS AND METHODS

1. Data Base

The generation of our database involved the following stages: (i) acquisition of digital chest radiograph, (ii) enhancement of nodule signal, and (iii) extraction of suspected nodule areas. [1]

(i) Acquisition of Digital Chest Radiographs - The PA (posterior-anterior) chest radiograph (14" × 17" actual size) was digitized to 2048 × 2500 × 10 bits by using Konica laser film scanner KDFR-S. For computational simplicity, the digitized chest films were miniaturized to 512 × 625 × 12 bits so that one pixel represented 0.7 mm × 0.7 mm. The chest radiographs could be lateral or straight and were selected mainly from routine cases at Georgetown University Medical Center.

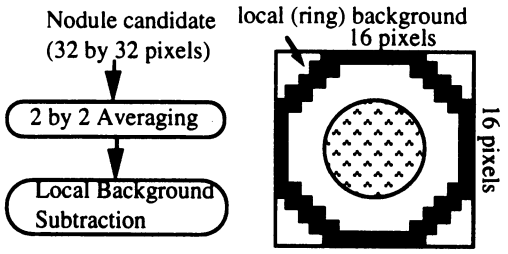
(ii) Enhancement of Nodule Signals - To enhance the nodule signals in the chest radiograph, we used an image subtraction technique [6] which subtracted a nodule-suppressed image (a median filtered image) from a nodule-enhanced image (a matched filtered image). The image subtraction method includes complex matrix multiplication, 2-D FFT, and inverse 2-D FFT.

(iii) Extraction of Suspected Nodule Areas - An area extraction procedure was conducted on the subtracted image to locate and isolate the suspected nodule areas. The procedure involved a contour searching process followed by the circularity test [1][6]. Each suspected nodule area was background corrected [6] and duplicated into a separate 32 by 32 pixel block as one nodule candidate. Since we concentrated on the early detection of small nodules, the block size of 32 × 32 pixels (about 22 mm × 22 mm) was sufficient to encompass the various sizes of small nodules (size smaller than 15 mm in diameter) in which we are interested.

The location and diagnoses of the suspected nodule areas were verified by an experienced radiologist of the Georgetown University Medical Center. Besides, the radiologist also assigned a linguistic label to each of the extracted nodule areas. These labels were "definitely a nodule", "probably a nodule", "possibly a nodule", "probably no nodule", and "definitely no nodule", and they were used to supervise the network learning (to be described in section 3.2.)

2. Preprocessing of Suspected Nodule Areas

The quality of chest radiographs varies because of the film development noise and the transmission properties of tissues, vessels, and ribs of different patients. In addition, because the lung field contains non-uniform anatomic structures, every nodule candidate has different local background noise. To remove the background noise, we used the image smoothing and local background subtracting methods (Figure 1(a)). First, the nodule candidate was smoothed by a 2 × 2 averaging filter and then its local background was subtracted from every pixel in the image block of suspected nodule area. The local background was defined as the average pixel value within the ring area as shown in Figure 1(b).



(a) Image pre-processing. (b) Average background.

Figure 1. Pre-processing of nodule candidates

3. The Two-Layer CNN

The two-layer convolution neural network (CNN) (Figure 2) is a simplified neocognitron model [13] without the C-layers. It consists a set of convolution kernels (the k × k links or synapses) which perform two-dimensional convolution on the image blocks of suspected nodule areas. The convolution kernels are to be trained by supervised learning and self-organized into a set of feature extractors. After training, each k by k convolution kernel functions as a feature detector and performs a specific feature extracting operation on the image blocks of the suspected nodule areas.

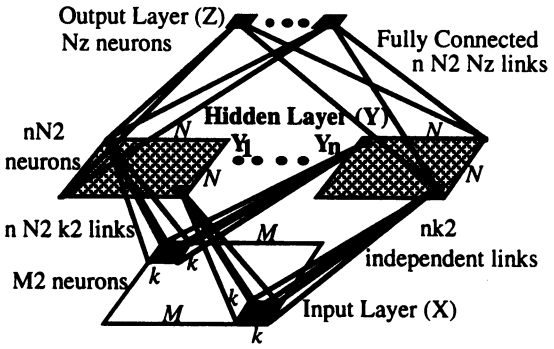


Figure 2. Architecture of a Two-Layer CNN

3.1 Network Architecture

The network has one input layer (X), one hidden layer (Y), and one output layer (Z). The input layer consists of M^2 neurons which corresponds to the M × M pixel pre-processed image block of nodule candidate.

The hidden layer Y is composed of n independent groups of feature planes (each has N × N neurons) which are designated by $Y_1, Y_2, \ldots, Y_n$. Each neuron in hidden layer Y takes input from a k × k

neighborhood on the input plane. For neurons in the same feature plane Y_j that are one neuron apart, their receptive fields (in the input layer) are one pixel apart. Moreover, neurons in the same feature map share the same set of k^2 weights (synaptic strength) and perform the same operation on the corresponding part of the input image. The total effect of the operation can be expressed as a two-dimensional convolution with a $k \times k$ kernel on the input image. All neurons in another feature map share another k^2 weights in the same way. Each hidden neuron in Y generates its output through an activation function (sigmoid function) and activation y_j of the neuron j in a feature map is given by

$$y_j(v, x, b) = \frac{1}{1 + \exp\{-[\sum_{i=1}^{k^2} (v_{ji}x_i) + b_j]\}} \quad (1)$$

where v_{ji} is the weight between hidden neuron j and input neuron i, k^2 is the convolution kernel size, and b_j is the neuron bias. Note that the $x_1,.., x_{k^2}$ are the parts of input image which are linked to the hidden neuron j.

The output layer is composed of N_Z neurons and is fully connected to hidden layer Y. The activation of the output neuron j is given by

$$z_j(w, y, g) = \frac{1}{1 + \exp\{-[\sum_{i=1}^{nN^2} (w_{ji}y_i) + g_j]\}} \quad (2)$$

where w_{ji} is the weight between output neuron j and hidden neuron i, nN^2 is the total number of neurons in the hidden layer, and g_j is the bias of output neuron j.

In summary, the network consists of $N_Z + M^2 + n \times N^2$ neurons (including input and output neurons) and $(N_Z + k^2) \times n \times N^2$ links (synapses) in which $n \times (N_Z \times N^2 + k^2)$ are independent links.

3.2 Fuzzification and Network Training

Currently, most radiologists are trained to make "crisp" decision, i.e., either "it's a nodule" or "no nodule". However, we believe that better network performance can be achieved if more diagnostic information is supplied during the network learning. Thus, we used the five linguistic labels (as described in (iii) of subsection 1 of section II) instead of "true" or "false" labels to supervise the network learning. The five labels were fuzzified (i.e., translated) into computer readable numbers and used as teaching signals during the network learning. Each label corresponds to one of the five output neurons in the output layer of the CNN. Therefore, we have five output neurons, instead of two neurons (true or false) [1][10][11][12]. Fuzzification of the five linguistic labels is described as follows.

A set of membership functions (Figure 3) was defined for the linguistic labels. The fuzzy membership functions were used to simulate the radiologist's diagnoses for the nodule candidates. Note that the membership functions of "probably a nodule" and "probably no nodule" are asymmetric about the firing output neurons 1 and 3, respectively.

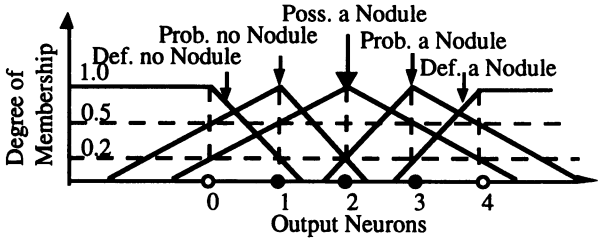


Figure 3. Fuzzy membership for five labels.

Network training is carried out by iteratively adjusting the synaptic strengths in the network so as to minimize the total error between the actual output state vector of the network and the target state vector (teaching signal). Training rule for updating the synaptic strengths can be obtained by taking gradient descent of a pre-specified error (objective) function with respect to the synaptic strengths [14]. In our simulation, the error function that is to be minimized by gradient descent is the fuzzy sum-of-squared error (FSSE) which is defined as

$$\text{FSSE} = \frac{1}{2} \sum_{p=1}^{N_p} \sum_{j=1}^{N_Z} (z_{p,j} - \varphi_{p,j})^2 . \quad (3)$$

where $z_{p,j}$ and $\varphi_{p,j}$ are the actual and desired activation value of output neuron j, respectively, when pattern p is present at the input of the network. The value of $\varphi_{p,j}$ represents the degree of "belongingness" of the input pattern p to jth class (output neuron). Both $z_{p,j}$ and $\varphi_{p,j}$ are real numbers between 0 and 1. The pattern index p runs over all the training patterns (i.e., the nodule candidates), N_p is the total number of patterns in the training set, and N_Z is the total number of output neurons.

During training, each output neuron was assigned with a membership value instead of choosing the single class (neuron) with the highest activation (delta function). This allows efficient modeling of ambiguous nodule candidates with appropriate weighting factors being assigned to the back-propagated errors depending upon the membership values at the corresponding outputs. The back-propagated "fuzzy" error was computed with respect to each desired output and network weights were adjusted by the back-propagation learning rule [14].

3.3 Defuzzification and Network Testing

After the network converged successfully to a minimum FSSE, the nodule candidates which were not used in training were applied as input to the

network. Note that the nodule candidates in the testing set were preprocessed in the same way as those in the training set before they were fed to the network. A defuzzification process [2] was employed to decipher the meaning of network outputs and to resolve the conflict between competing neurons. Several defuzzification techniques are available [2]. The defuzzification method we used is called centroid (or center-of-gravity) which takes the contribution of all fuzzy outputs and the degree of membership of each neuron into account. The defuzzified output value is called the system output activation (SOA) which is defined as

$$SOA = \left\{ \sum_{i=0}^{N_Z - 1} Z_i * (i - \frac{N_Z - 1}{2}) \right\} \Big/ \left\{ \sum_{i=0}^{N_Z - 1} Z_i \right\} \quad (3)$$

where Z_i is the fuzzy output of ith output neuron and N_Z is the number of output neurons. Higher value of SOA indicates highly suspected nodule area.

To test the network performance, first we organized the testing nodule image blocks into two subsets: one contained true nodules and the other contained false nodules. Then we computed SOA from the network outputs triggered by each image block of both subsets. Finally, the resulting two sets of SOA values were analyzed by the ROC method [3]. The A_Z (the area under the ROC curve) was used as performance index for evaluating the CNN network.

III. SIMULATION RESULTS

Experiment Setup - In our experiments, all the nodule candidates were generated by the pre-scan process operated at high sensitive mode [1]. There were 92 nodule image blocks (40 true nodules and 52 false nodules) in the training set and 554 nodule image blocks (472 false nodules and 82 true nodules) in the testing set. The radiographs which were pre-scanned for the testing set were not used in the pre-scan for the training set. Also, in the testing set, the multiple nodules were generated from the same film, but from different patients (A patient may have more than one chest radiograph). All nodule candidates, both in the training and testing set, were pre-processed in the same way as shown in Fig. 1.

Meanwhile, the sigmoid function was employed as the activation function for each neuron. Neither bias ($b_i = 0$ and $g_j = 0$, for $1 \le i \le nN^2$, $1 \le j \le N_Z$) nor momentum term ($\gamma = 0$) was used in the simulation. The weights of the network were initialized with small random numbers. The learning rate (η) was gradually decreased in discrete steps, taking values from the set {0.02, 0.01, 0.001, 0.0001}, until the network converged to a minimum FSSE at output neurons. During network learning, when a training pattern p was present at the network's input, each output neuron was assigned with a membership value (desired output value) according to the associated linguistic label (see Figure 3) of pattern p. The error between the desired output and actual output in the output layer (for all output neurons) was computed by using (3) and back-propagated for weight updating by Back-Propagation so that the weights were adjusted in proportion to their contribution to the error. The weights were updated after each presentation of a single training pattern.

To expand the limited training patterns and provide rotation-invariance information to the network, we employed the following training procedure. By rotating (at 90, 180, and 270 degrees) the training nodule candidate, we generated 3 additional image blocks. Moreover, by flipping over the original nodule candidate and rotating at 0, 90, 180, and 270 degrees, we got another 4 rotated version of nodule candidate. These 8 nodule candidates represented different viewing angle on the same suspected nodule area, they belong to the same "family" and shared the same diagnostic label. During the training phase, the whole "family" of nodule candidates were fed to the network, each family member was presented in random order to the network, and a feed-forward back-propagated pass was done for each one of them. All nodule candidates were pre-processed in the same way as shown in Figure 1. The CNN in our experiment failed to converge if no preprocessing (i.e., ring-background subtraction) was applied.

Results - We used CNN architecture consisting of 10 feature maps (hidden groups) and 5 output neurons. The ROC performance of 5-output neuron trained with delta function, and 5-output neuron trained with fuzzy function are shown in Figure 4. With 5 output neurons and use delta function as teaching signal during network learning, the performance index A_Z showed an average area of 0.78 under the ROC curve. The CNN with 5 output neurons trained by fuzzy membership value achieved mean Az of 0.84. The network performance is equivalent to 80% true-positive detection with 2 ~ 3 false-positive detections per chest image.

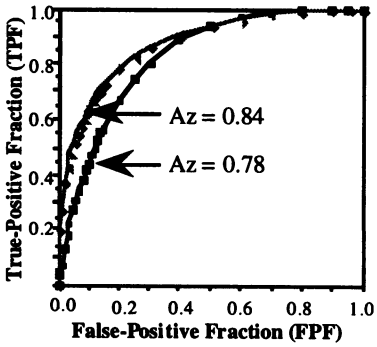


Figure 4. CNN trained with delta and fuzzy function.

IV. CONCLUSIONS AND DISCUSSION

From the computer simulations, we found that using fuzzy linguistic variables and membership functions was an effective method for network training. The fuzzy trained network is more robust in recognizing ambiguous image patterns. By investigating the false-positive cases, we found that the network was not shift invariant. Hence, more hidden layers and shifted training patterns should be used in order to enhance the CNN's ability of shift-invariance recognition [13].

The trained CNN was incorporated into our previous CAD scheme [1] to reduce the number of false-positive detections. The number of false-positive detections was reduced without eliminating any true nodule detection. It takes about 10 ~ 15 seconds (on DEC Alpha workstation) to capture the suspected nodule areas in one chest radiograph and make the final diagnosis.

The training speed of the CNN can be improved by adding momentum term or using other modifications of the learning algorithm [14][15]. The trained convolution kernels are to be analyzed so that the diagnostic information for identifying the true nodules can be found. Yet, network optimization and effective decision-making methods need to be developed to enhance the performance and processing speed of the CNN. Furthermore, the proposed fuzzy membership-based CNN method can also be applied to other diagnostic problems, for example, the detection of various abnormalities in chest radiographs and mammograms in clinical radiology.

References

[1]. Lin JS, Ligomenides PA, Freedman MT, and Mun SK. Application of Neural Networks for Improvement of Lung Nodule Detection in Radiographic Images. Symposium for Computer Assisted Radiology at Baltimore: 108 - 115, 1992.

[2]. Bezdek JC and Pal SK(Eds.). Fuzzy Models for Pattern Recognition. IEEE Press, NJ, 1992.

[3]. Swets JA and Pickett RM. Evaluation of Diagnostic Systems. Academic Press, New York, 1982.

[4]. Heelan RT, Flehinger BJ, Melamed MR, Zaman MB, Perchick WB, Caravelle JR, and Martini N. Non-Small-Cell Lung Cancer: Results of the New York Screening Program. Radiology: 151-289, 1984.

[5]. Ballard DH and Sklansky J. A Ladder-Structured Decision Tree for Recognizing Tumors in Chest Radiographs. IEEE Trans. on Computers 25 (5) 503-513, 1976.

[6]. Giger ML, Doi K, and MacMahon H. Image Feature Analysis and Computer-Aided Diagnosis in Digital Radiography 3 - Automated Detection of Nodules in Peripheral Lung Field Med Phy 15: 158-166, 1988.

[7]. Giger ML, Ahn N, Doi K, MacMahon H, and Metz CE. Computerized Detection of Pulmonary Nodules in Digital Chest Images: Use of Morphological Filters in Reducing False-Positive Detections. Med Phys 17: 861-865, 1990.

[8]. Matsumoto T, Yoshimura, Doi K, Giger ML, Kano A, MacMahon H, Abe K, and Montner SM. Image Feature Analysis of False-Positive Diagnoses Produced by Automated Detection of Lung Nodules. Investigative Radiology 8: 587-597, 1992.

[9]. Boone JM, Sigillita VG, and Shaber GS. Neural Networks in Radiology: An Introduction and Evaluation in a Signal Detection Task. Med Phy 17(2):234-241, 1990.

[10]. Wu YZ, Doi K, Giger ML, Metz CE, and Zhang W. Detection of Lung Nodules on Digital Radiographs: Comparison of Artificial Neural Networks and Discriminant Analysis. Radiological Society of North America, Physics (Digital image processing), Presentation #321, 1992.

[11]. Kim JH, Min BG, Han MC, and Lee CW. Computer-Assisted Detection of Lung Nodules by Using Artificial Neural Net. Radiological Society of North America, Physics (Digital image processing), presentation #325c, 1992.

[12]. Garg S and Floyd CE. Neural Network Localization of Pulmonary Nodule on Digital Chest Radiographs. RSNA'92, Physics (Digital image processing), Presentation #329, 1992.

[13]. Fukushima K and Miyaki S. Neocognitron: A New Algorithm for Pattern Recognition Tolerant of Deformations and Shifts in Position. Pattern Recognition 15(6):454-469, 1982.

[14]. Rumelhart DE, Hinton GE, and Williams RJ. Learning Internal Representations by Error Propagation. In Rumelhart & McClelland (Eds.), Parallel Distributed Processing: Explorations in the Micro Structure of Cognition. Vol. 1:Foundation. MIT Press, 1986.

[15]. Barmann F and Biegler-Konig F. On a Class of Efficient Learning Algorithms for Neural Networks. Neural Networks 5:139 - 144, 1989.